PRIVA'MOV: Analysing Human Mobility Through Multi-Sensor Datasets

Sonia Ben Mokhtar, Antoine Boutet, Louafi Bouzouina, Patrick Bonnel, Olivier Brette, Lionel Brunie, Mathieu Cunche, Stephane D 'Alu, Vincent Primault, Patrice Raveneau, et al.

PRIVA'MOV: Analysing Human Mobility Through Multi-Sensor Datasets

Sonia Ben Mokhtar, Antoine Boutet, Louafi Bouzouina, Patrick Bonnel, Olivier Brette, Lionel Brunie,
Mathieu Cunche, Stephane D'Alu, Vincent Primault, Patrice Raveneau, Herve Rivano, Razvan Stanica
University of Lyon, LIRIS, CNRS, INSA-Lyon, UMR5205, UMR5206 Triangle, UMR5600, INRIA, CITI, ENTPE, France
{sonia.benmokhtar, antoine.boutet, olivier.brette, lionel.brunie, stephane.dalu, mathieu.cunche, razvan.stanica, vincent.primault}
@insa-lyon.fr ; patrice.raveneau@univ-lr.fr ; herve.rivano@inria.fr ;
{louafi.bouzouina, patrick.bonnel}@entpe.fr

Abstract—**The wide adoption of mobile devices has created unprecedented opportunities to collect mobility traces and make them available for the research community to conduct interdisciplinary research. However, mobility traces available in the public domain are usually restricted to traces resulting from a single sensor (e.g., either GPS, GSM or WiFi). In this paper, we present the PRIVA'MOV dataset, a novel dataset collected in the city of Lyon, France on which user mobility has been collected using multiple sensors. More precisely, this dataset contains mobility traces of about 100 persons including university students, staff and their family members over 15 months collected through the GPS, WiFi, GSM, and accelerometer sensors. We provide in this paper both a quantitative and a preliminary qualitative analysis of this dataset. Specifically, we report the number of visited points of interests, GSM antennas and WiFi hotspots and their distribution across the various users. We finally analyse the uniqueness of human mobility by considering the various sensors.**

I. INTRODUCTION

The large adoption of mobile devices combined to their embedded localisation capabilities opens novel opportunities to provide mobility traces to the research community at large. Classical usages of mobility datasets include the analysis of user mobility patterns and their regularity [1], discovering hot places in a city [2] or studying privacy threats due to the disclosure of mobility data [3].

However most of the available datasets (e.g., the Cabspotting [4], the Geolife [5], or the T-Drive [6] datasets to named a few) contain data collections coming from only one sensor (i.e., the GPS, GSM or WiFi sensor). While single sensor mobility traces have been widely used in the literature to answer classical research questions, the availability of mobility data coming from multiples sensors opens the door for richer studies. This include comparative studies (where the data provided by a single sensor is compared to the data provided by another sensor) and compositional studies (where the data provided by a given sensor complements the data provided by another sensor). For instance, one may perform a comparative study of personal data leakage due to the sharing of the data provided by a given type of sensor versus the one due to the sharing of another type of sensor. On the other hand, one may combine the data provided by multiple sensors to increase the precision of user mobility data or to compensate for the lack of data from one sensor. For instance, one may use the WiFi and GSM mobility data to remove erroneous localisations provided by the GPS sensor.

We present in this paper the PRIVA'MOV dataset that contains the mobility traces of 100 users around the city of Lyon collected using various sensors, namely the GPS, WiFi, GSM, and accelerometer sensors. This dataset has been collected from October 2014 to January 2016 by equipping volunteer students from three universities, staff members and sometimes their relatives with smartphones on which a crowdsensing application was periodically collecting records from the above mentioned sensors.

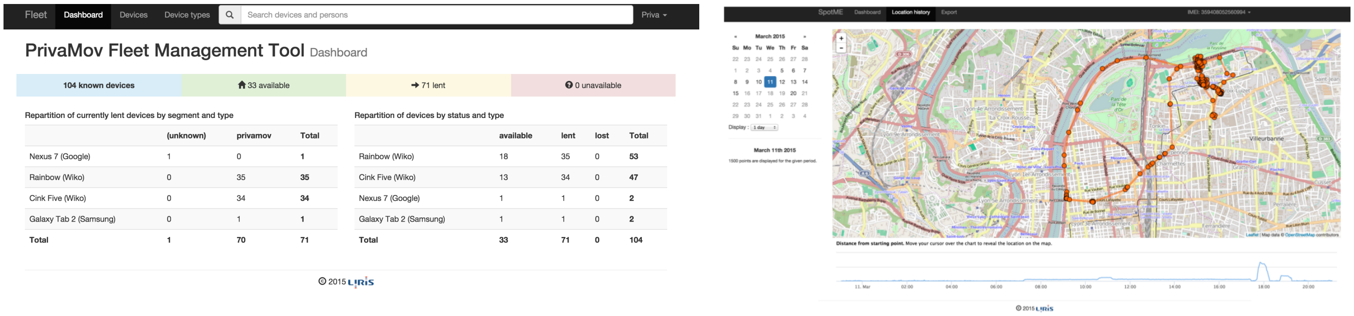
In the remaining of this paper we present the data collection process (Section II), a quantitative analysis of the resulting dataset (Section III) as well as a preliminary qualitative analysis of its records (Section IV). More precisely, we analyse the relation between meaningful locations of the city and users and the uniqueness of mobility traces of users considering the various sensors. We finally draw our conclusions and future research directions (Section V).

II. THE PRIVA'MOV CROWDSENSING CAMPAIGN

The PRIVA'MOV dataset has been collected during a crowdsensing campaign that took place in the city of Lyon from October 2014 to January 2016 in the context of PRIVA'MOV project funded by the LABEX IMU[1] funding agency. In the context of this project 100 smartphones (52 Wiko Rainbow and 48 Wiko Cink 5) have been equipped with a crowdsensing application and distributed to students, staff members and their relatives from three universities: INSA Lyon, ENS Lyon and Université Claude Bernard, Lyon 1. Volunteers were asked to use the PRIVA'MOV phone as their primary phone and to carry it during their daily activities. The crowdsensing application was a modified version of the application developed in the Funf project [2]. A complementary fleet management web application and a trace visualisation tool depicted in Figure 1a and Figure 1b, respectively, have also been developed.

The crowdsensing application has been configured to collect the data every time the system used them (e.g., change of location, new WiFi scan). In order to save battery, the collected data were uploaded to the server only when the smartphone was connected to a WiFi network. The resulting dataset is described in the following section.

[1]LABEX IMU: http://imu.universite-lyon.fr
[2]Funf: http://funf.org



(a) Fleet management portal (b) Visualisation tool

Fig. 1. PRIVA'MOV Fleet management portal (Figure 1a) ; PRIVA'MOV trace visualisation tool (Figure 1b).

III. PRIVA'MOV QUANTITATIVE ANALYSIS

We describe in this section a quantitative analysis of the PRIVA'MOV dataset. Table I shows the number of records collected by the various sensors in the overall dataset.

Sensor	Number of Records
WiFi	25,655,480
Cellular	8,076,512
GPS	156,041,576
Accelerometer	90,066,831
Battery	7,008,504

TABLE I. THE PRIVA'MOV DATASET CONTAINS MOBILITY DATA COLLECTIONS CAPTURED THROUGH DIFFERENT SENSORS.

From these data collections, it is then possible to extract mobility traces. We call a mobility trace a list of spatio-temporal points belonging to a given user. This can be done in the WiFi and GSM data collections by associating a GPS location to each WiFi and GSM antenna by relying on public datasets such as WiGLE [3] or Google [4].

In the PRIVA'MOV dataset we did not perform the mapping of WiFi and GSM antennas to GPS locations. Instead, we use the unique identifier of the cellular antenna to which the user is connected (respectively the MAC address of the WiFi access point to which the user is attached, or discovered through a periodic WiFi scan) as a spatial indicator of the user location.

In addition, mobility traces built from the GPS data collections could be enriched with user Points of Interest (POIs). A POI is a meaningful location where the user has marked a significant stop. To compute POIs, we used a methodology similar to [7]. The idea behind this method is to identify restricted areas where users stay more than a specific duration. More precisely, POIs can be extracted using a simple spatio-temporal clustering algorithm parametrised with a maximum POI diameter d and a minimum stay time t. This POIs extraction is done in two clustering steps, the first one identifies POIs for each user and the second one assigns identifiers to unique POIs, thus allowing to identify POIs shared by several users. For instance, Figure 2 illustrates the POIs of users in the Lyon sub-area for a diameter of 250 meters ($d = 250$) and a stay time of 30 minutes ($t = 30$).

[3]WiGLE: Wireless Network Mapping, http://wigle.net

[4]Google Maps Geolocation API: https://developers.google.com/maps

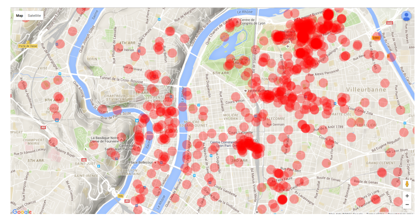
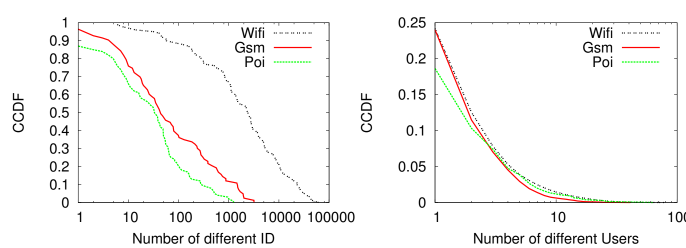


Fig. 2. Location of Points Of Interests (POIs) across the Lyon area.

In the PRIVA'MOV dataset, we extracted from the GPS data collections the set of all POIs and assigned an identifier to each unique POI inside the whole dataset. Then, similarly to the GSM and WiFi datasets we used unique POI identifiers as a spatial component.

Let us consider the extracted mobility traces as explained above (i.e., using cellular antenna IDs, WiFi mac addresses, and POI IDs from the data collections). To analyse the resulting mobility traces, Figure 3 shows the Complementary Cumulative Distribution Function (CCDF, defined as $P(X > x)$). Figures 3a shows the number of unique GSM antennas, WiFi access points and POIs per user. Figures 3b depicts the number of unique users identified per GSM antenna, WiFi access point and POI. These tail distributions show that most of GSM antennas, WiFi access points and POIs have been visited only by one user. For instance, 75% of the WiFi access points have been seen by only one user. Conversely, mobility traces of most of the users are composed of several GSM antennas, WiFi access points and POIs. Finally, these figures show that users have discovered more WiFi access points than GSM antennas, and the number of POIs is lower than the two others.



(a) # of Identifiers (b) # of Users

Fig. 3. The tail distributions of the number of different GSM antennas, WiFi access points, and POI per user (Figure 3a) ; and the number of unique users per GSM antenna, WiFi access point, and POI (Figure 3b).

IV. PRIVA'MOV QUALITATIVE ANALYSIS

In this section, we report two qualitative experiments conducted on the PRIVA'MOV dataset. Specifically, we first analyse the spatial relationship between POIs inferred from the activity of users. Then we study the uniqueness of user mobility traces over the three data types.

A. POI relationship inference

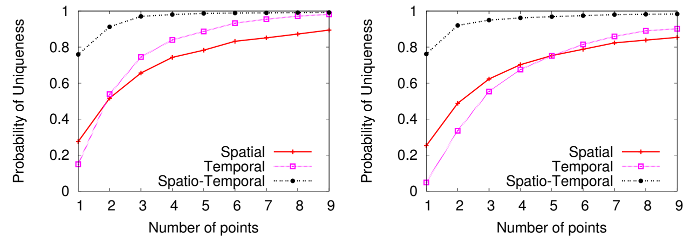
In this experiment, we analysed the relationship between POIs appearing in the dataset. Specifically, we consider that two POIs are related if there exist at least one user that visited the two POIs. Figure 4 shows the resulting graph obtained by extracting the relationship between all the POIs of the PRIVA'MOV dataset. From this figure, we observe that while some POIs have been widely visited by many users (at the bottom left), some places have been only visited by sub groups of users (clusters of point at the top right).

Fig. 4. Relationship between points of interest inferred from the mobility of users (two points of interest are connected if at least one user has visited these places).

B. Uniqueness of human mobility

To quantify the uniqueness of mobility traces, we use the methodology proposed by De Montjoye and all in [8]. More precisely, for each mobility trace T, we evaluate the uniqueness of a given sub-trace I_p of p randomly chosen spatio-temporal points. A sub-trace I_p is said to be unique if only one user has $I_p \in T$. To measure this uniqueness, we performed a brute force search of users who have the p points composing I_p in their mobility trace T. The size of this set of users sharing the same I_p, noted k, characterizes the uniqueness of the sub-trace I_p. If $k = 1$, the sub-trace is unique. The uniqueness of traces is estimated as the percentage of 2500 random sub-traces that are unique given the p points composing them. We use the same methodology to evaluate the uniqueness of spatial or temporal only mobility traces. In this case, the sub-trace I_p contains spatial or temporal points, respectively.

We report the uniqueness of mobility traces built from the GPS and the WiFi traces. Figure 5 depicts the probability to be unique according to the number of points in the considered sub-trace. The evaluation reports results for spatial, temporal, and spatio-temporal traces. As shown in these figures, the



(a) GSM (b) WiFi

Fig. 5. Four spatio-temporal points are enough to uniquely identify 97% of the individuals.

results for spatio-temporal mobility traces from GSM and WiFi show a strong uniqueness. More precisely, four spatio-temporal points are enough to identify on average 97% of the users. This high uniqueness is the result of combining the temporal and the spatial mobility information of users, which are discriminative enough to uniquely identify them. Although this analysis uses a smaller set of data, this result comforts and generalizes the previous study performed in [8] on call logs.

V. CONCLUSIONS

To address the lack of multi-sensor mobility datasets, the project PRIVA'MOV developed and deployed a crowdsensing platform to collect mobility traces from a sample of real users equipped with mobile phones. We presented the resulting dataset and reported quantitative and qualitative experiments over this dataset, which show the potential of using the latter for answering novel research questions. The PRIVA'MOV datasets is available for research community upon request and under a set of usage conditions. Our future work targets the data collection of multi-sensor information over a larger population of users with a real time data analysis and a validation tool to ask the users to validate or not the inferred information (i.e., leaked personal information, mobility behavior).

REFERENCES

[1] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, June 2008.

[2] J. Yuan, Y. Zheng, and X. Xie, "Discovering regions of different functions in a city using human mobility and pois," in *KDD*, 2012, pp. 186–194.

[3] S. Gambs, M.-O. Killijian, and M. N. n. del Prado Cortez, "Show me how you move and i will tell you who you are," in *SPRINGL*, New York, NY, USA, 2010, pp. 34–41.

[4] M. Piorkowski, N. Sarafijanovic-Djukic, and M. Grossglauser, "CRAWDAD dataset epfl/mobility (v. 2009-02-24)," Downloaded from http://crawdad.org/epfl/mobility/20090224, Feb. 2009.

[5] Y. Zheng, L. Zhang, X. Xie, and W.-Y. Ma, "Mining Interesting Locations and Travel Sequences from GPS Trajectories," in *WWW*. ACM, 2009, pp. 791–800.

[6] J. Yuan, Y. Zheng, X. Xie, and G. Sun, "Driving with Knowledge from the Physical World," in *KDD*, 2011, pp. 316–324.

[7] Y. Zheng, L. Liu, L. Wang, and X. Xie, "Learning transportation mode from raw gps data for geographic applications on the web," in *WWW*, 2008, pp. 247–256.

[8] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the crowd: The privacy bounds of human mobility," *Scientific Reports*, vol. 3, 2013.